ORDERING DISTRIBUTIONS ON THE CIRCLE WITH RESPECT TO UNIFORMITY

BY

FRED W. HUFFER

TECHNICAL REPORT NO. 406
JUNE 29, 1988

Prepared Under Contract
N00014-86-K-0156 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. Introduction

Let $\mathbf{C}$ be a circle with circumference equal to one. (For notational convenience, we frequently identify $\mathbf{C}$ with an interval of length one on the real line.) For probability measures on $\mathbf{C}$ we will define orderings $\ll$ such that $\mu_1 \ll \mu_2$ means roughly that μ_1 is "more uniform" than μ_2 or that μ_2 is "more clumped" than μ_1. Distributional orderings on the real line $\mathbf{R}$ have proven to be very useful in many areas of probability and statistics and it is hoped that orderings on $\mathbf{C}$ will also prove to be of some use. Later in this paper we shall present applications of the orderings $\ll$ to the following areas: random coverage problems on the circle, the distribution of spacings, and the analysis of directional data. In this section we shall motivate and define the orderings and present some of their basic properties.

To define the orderings we need the following notation. Let U be a random variable uniformly distributed on $\mathbf{C}$. A random arc on $\mathbf{C}$ having length b and clockwise endpoint U will be denoted $[U,U+b)$. For any probability measure μ on $\mathbf{C}$ and length $b \in (0,1)$ we define the random variable $D(\mu,b) = \mu([U,U+b))$. Please note that throughout the paper we shall use μ (perhaps with a subscript or superscript) to denote a probability measure on $\mathbf{C}$.

Our orderings have the form $\mu_1 \ll \mu_2$ if and only if $D(\mu_1,b)$ is "less variable" than $D(\mu_2,b)$ for all $b \in (0,1)$. To motivate this type of definition we note first that

(1.1) $\quad ED(\mu,b) = b$ for all μ and all $b \in (0,1)$.

(See Section 5 for a proof of this; all proofs are given in Section 5.) Now consider the two extreme cases: let λ be the uniform distribution on $\mathbf{C}$ and δ_x be the degenerate distribution which assigns all its mass to an arbitrary point $x \in \mathbf{C}$. Clearly $D(\lambda,b) = b$ with probability one and

$$D(\delta_x,b) = \begin{cases} 0 & \text{with probability } 1-b, \\ 1 & \text{with probability } b. \end{cases}$$

For distributions on $[0,1]$ with mean b, these are the two extremes with respect to variability. This suggests using the "variability" of $D(\mu,b)$ to measure the "uniformity" of μ.

Using two different notions of "variability" we obtain the orderings $\underset{v}{\ll}$ and $\underset{c}{\ll}$ defined as follows:

$$\mu_1 \underset{v}{\ll} \mu_2 \text{ if } \mathrm{Var}(D(\mu_1,b)) \leq \mathrm{Var}(D(\mu_2,b)) \text{ for all } b\in(0,1).$$

$$\mu_1 \underset{c}{\ll} \mu_2 \text{ if } E\Psi(D(\mu_1,b)) \leq E\Psi(D(\mu_2,b)) \text{ for all convex functions } \Psi \text{ and all } b\in(0,1).$$

The letters v and c have been chosen to remind us of the words "variance" and "convex". Using (1.1) and the fact that $\Psi(x)=x^2$ is convex we obtain

(1.2) $\mu_1 \underset{c}{\ll} \mu_2$ implies $\mu_1 \underset{v}{\ll} \mu_2$

In order to vindicate the above definitions, we shall now list some basic properties of the orderings. These properties are, for the most part, ones that any reasonable "uniformity" ordering ought to possess. The properties hold for both $\underset{v}{\ll}$ and $\underset{c}{\ll}$, so they are stated without the subscripts v and c.

(1.3) $\lambda \ll \mu \ll \delta_x$ for all μ. Here λ is the uniform distribution and δ_x is a point mass at x.

(1.4) $\mu_1 \ll \mu_2$ and $\mu_2 \ll \mu_3$ imply $\mu_1 \ll \mu_3$.

Convolution of measures on C can be defined as follows. Identify C with the interval [0,1). If X and Y are random variables with probability measures μ and ξ respectively, then the convolution $\mu*\xi$ is the probability measure of (X+Y) mod1 . Intuitively speaking, the operation of convolution smooths distributions and makes them more uniform, so that it is natural to have

(1.5) $\mu*\xi \ll \mu$ for all probability measures μ and ξ on C.

In this paragraph, we shall identify C with the interval [-1/2, 1/2). A probability density f defined on C is said to be symmetric and unimodal if $f(x)=f(-x)$ for all x and f is nonincreasing on the interval [0,1/2). Let μ_1 and μ_2 have densities f_1 and f_2 respectively. If f_1 and f_2 are symmetric and unimodal, and f_2 is "more peaked" about zero than f_1, then it seems reasonable to say that μ_1 is more uniform than μ_2, and we can indeed show the following:

(1.6) If f_1 is symmetric and unimodal and

$$\int_{-x}^{x} f_1(u)du \leq \int_{-x}^{x} f_2(u)du \text{ for } 0 < x < 1/2, \text{ then } \mu_1 \ll \mu_2.$$

(It turns out that we do not need either symmetry or unimodality for f_2, but this is not very intuitive.) Here, we are using "more peaked" in the same sense as Birnbaum (1948).

Finally, we note two convexity properties. The quantities $p_1, p_2, \ldots, p_n$ are nonnegative and sum to one.

(1.7) If $\mu_i \ll \mu$ for all i, then $\sum_{i=1}^{n} p_i\mu_i \ll \mu$.

(1.8) If $\mu \ll (1-\beta)\mu + \beta\mu_i$ for $0 < \beta \leq 1$ and all i, then $\mu \ll \sum_{i=1}^{n} p_i\mu_i$.

2. Ordering Discrete Distributions

In this section we give explicit conditions which allow us to verify whether or not the orderings $\underset{c}{\ll}$ and $\underset{v}{\ll}$ hold for some simple types of discrete distributions. In particular, these conditions allow us to compare empirical distributions so that, given two samples each containing n observations on $\mathbb{C}$, we can determine if one sample is more uniformly dispersed than the other. For future convenience we define $\mathcal{S}^n$ to be the set of vectors $(x_1, x_2, \ldots, x_n)$ satisfying $\sum_{i=1}^{n} x_i = 1$ and $x_i \geq 0$ for all i.

Suppose a measure μ has n atoms each having a mass of $1/n$, that is, μ is the empirical distribution of a sample of n observations. Then μ is determined (up to a rotation) by the vector $x = (x_1, x_2, \ldots, x_n)$ of spacings between consecutive atoms and we shall write $\mu = SP(x)$. More precisely, for any vector $x \in \mathcal{S}^n$, we define the probability measure SP(x) on $\mathbb{C}$ (identified with $[0,1)$) by

$$SP(x) = \frac{1}{n}\sum_{i=0}^{n-1} \delta_{u_i} \text{ where } u_0 = 0 \text{ and } u_i = \sum_{k=1}^{i} x_k \text{ for } 1 \leq i \leq n-1.$$

Now consider a measure μ having n atoms (of varying masses) equally spaced around the circle $\mathbb{C}$ so that consecutive atoms are separated by the distance $1/n$. This measure μ is determined (up to a rotation) by the vector $x = (x_1, x_2, \ldots, x_n)$ of probability masses for the consecutive atoms and we shall write $\mu = PM(x)$. More precisely, for any vector $x \in S^n$, we shall define the probability measure $PM(x)$ on $\mathbb{C}$ by

$$PM(x) = \sum_{i=0}^{n-1} x_{i+1} \delta_{i/n} .$$

The names SP and PM are intended to remind us of "spacings" and "probability masses" respectively. The measures $SP(x)$ and $PM(x)$ are inverses or duals in a sense to be developed in Section 3 so that by the result (3.1) we obtain

(2.1) $\quad SP(x) \underset{c}{\ll} SP(y)$ if and only if $PM(x) \underset{c}{\ll} PM(y)$.

We now present a condition on x and y which is necessary and *sufficient for* $SP(x) \underset{c}{\ll} SP(y)$. This condition is stated in terms of the majorization ordering for vectors which is treated at length in the books by Hardy, Littlewood and Polya (1952) and Marshall and Olkin (1979). For any vector $(x_1, x_2, \ldots, x_n)$, let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ denote the coordinates arranged in increasing order $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$. If two vectors x and y in $\mathbb{R}^n$ satisfy

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \text{ and } \sum_{i=k}^{n} x_{(i)} \le \sum_{i=k}^{n} y_{(i)} \text{ for } 2 \le k \le n ,$$

we say that x is majorized by y and write $x \underset{m}{\le} y$.

For any vector $x = (x_1, x_2, \ldots, x_n) \in S^n$ and integer k with $0 \le k \le n$ we define the vector $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})$ by

$$x_i^{(k)} = \sum_{j=0}^{k-1} x_{i+j} \quad \text{for all } i.$$

Here the coordinates of x are considered to be arranged in a circle so that $x_{n+i} = x_i$ for all i. Note that $x^{(1)} = x$, $x^{(0)} = (0,0,\ldots,0)$ and $x^{(n)} = (1,1,\ldots,1)$. Our condition can now be stated. For x and y in S^n,

(2.2) $SP(x) \underset{c}{\ll} SP(y)$ if and only if $x^{(k)} \underset{m}{\leq} y^{(k)}$ for $1 \leq k \leq n-1$.

The relationship (2.1) becomes false upon replacing $\underset{c}{\ll}$ by $\underset{v}{\ll}$, so we must state separate conditions for $SP(x) \underset{v}{\ll} SP(y)$ and $PM(x) \underset{v}{\ll} PM(y)$. These are now given. For x and y in S^n,

(2.3) $PM(x) \underset{v}{\ll} PM(y)$ if and only if

$$\sum_{i=1}^{n}(x_i^{(k)})^2 \leq \sum_{i=1}^{n}(y_i^{(k)})^2 \text{ for } 1 \leq k \leq n-1, \quad \text{and}$$

(2.4) $SP(x) \underset{v}{\ll} SP(y)$ if and only if

$$(x^{(1)},x^{(2)},\ldots,x^{(n-1)}) \underset{m}{\leq} (y^{(1)},y^{(2)},\ldots,y^{(n-1)}).$$

Here $(x^{(1)},x^{(2)},\ldots,x^{(n-1)})$ denotes the vector obtained by joining (stringing together) the vectors $x^{(1)},x^{(2)},\ldots,x^{(n-1)}$ into one long vector having $n(n-1)$ components.

It is straightforward to check the conditions (2.2), (2.3) or (2.4) for any given x and y in S^n, so these conditions are useful in constructing counterexamples like the following. Let $x = (.3,.3,.2,.2)$ and $y = (.55,0,.45,0)$. We can easily verify that $SP(x) \underset{v}{\ll} SP(y)$ is true, but that both $SP(x) \underset{c}{\ll} SP(y)$ and $PM(x) \underset{v}{\ll} PM(y)$ are false. This shows that the orderings $\underset{c}{\ll}$ and $\underset{v}{\ll}$ are in fact different and also that the relation (2.1) becomes false upon replacing $\underset{c}{\ll}$ by $\underset{v}{\ll}$. Similarly, with $s = (.4,.1,.4,.1)$ and $t = (.55,.15,.15,.15)$, we find that $PM(s) \underset{v}{\ll} PM(t)$ is true, but that $SP(s) \underset{v}{\ll} SP(t)$ is false.

We wish to note one more counterexample. Let

$$\mu_1 = SP(\tfrac{1}{12}(1,2,3,2,1,3)) \text{ and } \mu_2 = SP(\tfrac{1}{12}(1,2,3,1,3,2)).$$

It is easy to verify that both $\mu_1 \underset{c}{\ll} \mu_2$ and $\mu_2 \underset{c}{\ll} \mu_1$ are true. The measures μ_1 and μ_2 are not equivalent in any obvious sense; neither measure can be made to coincide with the other by using rotations or reflections of the circle. Thus $\underset{c}{\ll}$ is not a legitimate partial ordering; it satisfies the transitive property (1.4), but not the anti-symmetry property.

3. The Preservation of Orderings by Inverses

For any measure μ we define an inverse (or dual) measure μ^* as follows. Identify $\mathbb{C}$ with $[0,1)$ and let the random variable U be uniform on $[0,1)$. The random variable $X=\mu([0,U])$ can be regarded as a random point on $\mathbb{C}$ so long as 0 and 1 are taken to represent the same point on $\mathbb{C}$. Define μ^* to be the probability measure of X. If μ has no atoms and has support on all of $\mathbb{C}$, then the function $F(x) = \mu([0,x])$ is invertible and $F^{-1}(x) = \mu^*([0,x])$ for $0 \le x < 1$. The measure μ^* depends on the arbitrary choice of the origin (zero point) on $\mathbb{C}$, but in a trivial way. Different choices of the origin lead to measures μ^* which differ only by a rotation.

If the measures μ_1 and μ_2 are identical or can be made so by rotating one of the measures, we shall write $\mu_1 = \mu_2$. For the discrete measures in Section 2 it is easily seen that $\mu = SP(y)$ implies $\mu^* = PM(y)$ and similarly, $\mu = PM(y)$ implies $\mu^* = SP(y)$. In general, $\mu^{**} = \mu$ for all μ.

It will be convenient to use the following notation. Let X and Y be bounded random variables with distributions F and G respectively. If $E\Psi(X) \le E\Psi(Y)$ for all convex functions Ψ, we say that $X \lesssim_{H} Y$ or equivalently $F \lesssim_{H} G$. This type of ordering is well known and has many uses in queueing theory (see Stoyan(1983)). The ordering $\lesssim_{H}$ is closely related to the majorization ordering $\lesssim_{m}$ for vectors, hence the similarity in the notation.

The main result of this section is that inversion preserves the ordering $\ll_{c}$, that is,

(3.1) $\quad \mu_1 \ll_{c} \mu_2$ if and only if $\mu_1^* \ll_{c} \mu_2^*$.

If we restrict ourselves to measures μ which have no atoms and have support on all of $\mathbb{C}$, we can restate (3.1) in a more concrete fashion. Under this restriction, for any $y \in \mathbb{C}$ and $t \in (0,1)$ we can define $L(y,\mu,t)$ to be the length of the unique arc A on $\mathbb{C}$ which has clockwise endpoint y and satisfies $\mu(A)=t$. Let Y_i be distributed according to μ_i for $i=1,2$. Then

(3.2) $\quad \mu_1 \ll_{c} \mu_2$ if and only if $L(Y_1,\mu_1,t) \lesssim_{H} L(Y_2,\mu_2,t)$ for all $t \in (0,1)$.

The relation (3.1) becomes false upon replacing $\ll_{c}$ by $\ll_{v}$. See the discrete counterexamples at the end of section 2. However, something

resembling (3.2) can be shown to hold for the ordering $\ll_v$. For any points $x \neq y$ on $\mathbb{C}$, define $d(x,y)$ to be the length of the arc $[x,y]$ having clockwise endpoint x and counterclockwise endpoint y. When $x=y$, we randomize and set $d(x,y)=0$ or 1 with each value having probability $1/2$. Note that $d(x,y)+d(y,x)=1$ unless $x=y$. Let X_1, Y_1 be independent with distribution μ_1 and X_2, Y_2 be independent with distribution μ_2. We can show that

(3.3) $\mu_1 \ll_v \mu_2$ if and only if $d(X_1,Y_1) \leq_{st} d(X_2,Y_2)$.

Specializing this to discrete measures of the form $SP(x)$ leads to the result given in (2.4).

4. Applications

This section contains some applications of the uniformity orderings $\ll_c$ and $\ll_v$. In the first two applications the orderings are used to give precise versions of some fairly intuitive qualitative statements.

Suppose n arcs having arbitrary lengths $b_1, b_2, \ldots, b_n$ are placed uniformly and independently on $\mathbb{C}$. Let $\mathcal{A}$ denote the covered region,

$$\mathcal{A} = \bigcup_{i=1}^{n} [U_i, U_i + b_i)$$

where $U_1, U_2, \ldots, U_n$ are independent and uniformly distributed on $\mathbb{C}$. Intuition suggests that if μ_1 is "more uniform" than μ_2, then the covered mass $\mu_1(\mathcal{A})$ will be "less variable" than $\mu_2(\mathcal{A})$. We can indeed show that

(4.1) If $\mu_1 \ll_v \mu_2$, then $\mathrm{Var}(\mu_1(\mathcal{A})) \leq \mathrm{Var}(\mu_2(\mathcal{A}))$.

Note that when $n=1$ this statement follows immediately from the definition of $\ll_v$. Another result along the lines of (4.1) is

(4.2) If $\mu_1 = \mu_2 * \mu_3$ for some measure μ_3, then $\mu_1(\mathcal{A}) \leq_{st} \mu_2(\mathcal{A})$.

Remember that $*$ denotes convolution. The condition $\mu_1 = \mu_2 * \mu_3$ is much stronger than $\mu_1 \ll_v \mu_2$ (see result (1.5)).

The second application concerns the distribution of spacings and higher order spacings. Let $X_1, X_2, \ldots, X_n$ be i.i.d. from the distribution μ_1. Suppose $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are the points X_i arranged in counterclockwise order beginning with $X_{(1)} = X_1$. These points divide $\mathcal{C}$ into n segments whose lengths (called spacings) are denoted $S_1, S_2, \ldots, S_n$. To be precise we take $S_i = d(X_{(i)}, X_{(i+1)})$ with d as in (3.3) and $X_{(n+1)} = X_{(1)}$. Similarly, let $T_1, T_2, \ldots, T_n$ be the spacings between n i.i.d. points from the distribution μ_2.

The "higher order spacings" are simply sums of consecutive spacings. The properties of spacings and higher order spacings have been extensively studied. For information and further references on this area see Pyke (1965), Holst (1979) and Guttorp and Lockhart (1988).

When μ_1 is "more uniform" than μ_2, it seems intuitive that the spacings (and higher order spacings) under μ_1 will be "less variable" than those under μ_2. We can prove that

(4.3) If $\mu_1 \underset{c}{\ll} \mu_2$, then $\sum_{i=1}^{k} S_i \underset{M}{\leq} \sum_{i=1}^{k} T_i$ for $1 \leq k \leq n-1$.

Our final area of application is the analysis of directional data. A basic problem in this area is that of testing for uniformity: Does a given sample of n observations look like it was obtained by random sampling from the uniform distribution on $\mathcal{C}$? If the alternatives to uniformity have no preferred direction, it is reasonable to use test procedures which are invariant under rotations of the circle. Such an invariant procedure can be described in terms of a test statistic $\phi(S_1, S_2, \ldots, S_n)$ which is a function of the consecutive spacings $S_1, S_2, \ldots, S_n$ between the n data points; we reject the hypothesis of uniformity when ϕ is sufficiently large. The book by Mardia (1972) reviews many of the test statistics which have been proposed for testing uniformity.

What functions ϕ will lead to reasonable tests for uniformity? Functions which are increasing with respect to $\underset{c}{\ll}$ or $\underset{v}{\ll}$ are obvious candidates. Let x and y belong to $\mathcal{S}^n$. A function ϕ will be called *increasing* if $\phi(x) \leq \phi(y)$ whenever $SP(x) \ll SP(y)$. A prefix of *c* or *v* will indicate whether we refer to $\underset{c}{\ll}$ or $\underset{v}{\ll}$. Clearly, any *v*-increasing function is also *c*-increasing. Using the results (2.2) and (2.4) we can easily show the following:

(4.4) If Φ is a Schur-convex function of n arguments, then $\phi_k(x) = \Phi(x^{(k)})$ is a *c*-increasing function for $1 \leq k \leq n-1$.

(4.5) If Φ is a Schur-convex function of $n(n-1)$ arguments, then $\phi(x) = \Phi(x^{(1)},x^{(2)},\ldots,x^{(n-1)})$ is a ν-increasing function.

See Marshall and Olkin (1979) for information on Schur-convex functions. Additional c-increasing functions may be obtained by taking various combinations of the functions in (4.4); If $\psi_1, \psi_2, \ldots, \psi_m$ are c-increasing functions, then $\psi_1+\psi_2+\ldots+\psi_m$ and $\max(\psi_1, \psi_2, \ldots, \psi_m)$ are c-increasing, etc.

We note that many of the proposed test statistics are indeed either c or ν-increasing. We shall use the nomenclature of chapter 7 of Mardia (1972). This source contains definitions for all the statistics mentioned below except for the scan test which is described in Naus (1966) and Cressie (1977).

The following statistics are all c-increasing: Kuiper's test, the range test, the equal spacings test, Hodges-Ajne's test and the scan test. These statistics are all closely related so that in section 5 we shall discuss only two of them: the scan test and Kuiper's test.

Cressie (1976, 1979) considers test statistics of the form

$$\phi(x) = \sum_{i=1}^{n} h(x_i^{(k)})$$

with emphasis on the cases $h(u) = -\log u$ and $h(u) = u^2$. According to (4.4), a statistic of this form will be c-increasing whenever h is convex.

Examples of ν-increasing statistics can be found in the general class of tests for uniformity developed by Beran (1968, 1969). These test statistics can all be rewritten in the form

$$B_n(x) = k + \sum_{i=1}^{n(n-1)} h(y_i) \quad \text{for some constant } k \text{ and function } h,$$

where y_i are the components of the vector $(x^{(1)},x^{(2)},\ldots,x^{(n-1)})$.

According to (4.5), the function B_n will be ν-increasing whenever h is convex. Using this result we can show that both Watson's U_n^2 test and Ajne's A_n test are ν-increasing. See the discussion on pages 190-192 of Mardia's book.

The Rayleigh test is an example of a test statistic which is <u>not</u> c-increasing; counterexamples are easy to construct. This is to be expected because the Rayleigh test is intended only for use in detecting unimodal alternatives to uniformity and is not sensitive to many other forms of nonuniformity.

5. Proofs

Proof of (1.1). We note that

$$E\, D(\mu,b) = \int(\int I_{[x,x+b)}d\mu)dx$$

where $I_{[x,x+b)}$ is the indicator function of the given interval and both integrations are over all of $\mathbb{C}$ which is identified with $[0,1)$ using addition modulo one. Interchanging the order of integration yields (1.1).

Proof of (1.2) is immediate.

Proof of (1.3). Because of (1.2), it suffices to prove (1.3) for the ordering $\ll_c$. In terms of the ordering $\lesssim_M$ introduced in Section 3, we must show that $D(\lambda,b) \lesssim_M D(\mu,b) \lesssim_M D(\delta_x,b)$ for all μ and b. The extremal distributions of the ordering $\lesssim_M$ are well known (see Example 1.9(b), page 25 of Stoyan (1983)) so that this result follows from (1.1) and the accompanying discussion.

Proof of (1.4) is immediate.

Proof of (1.5). Again, by (1.2) it suffices to consider only $\ll_c$. To simplify the presentation we carry out the proof only for measures ξ which are discrete with finitely many atoms. In this case (1.5) follows easily from (1.7) which is proved later. Suppose ξ has atoms of mass p_i at the points x_i for $1 \le i \le n$. Define the measure μ_i by $\mu_i(B + x_i) = \mu(B)$ for all measurable B contained in $\mathbb{C}$. Then $\mu * \xi = \sum_i p_i\mu_i$ and $\mu_i \ll_c \mu$ for all i. Now using (1.7) completes the proof.

Proof of (1.6). Again, it suffices to consider only $\ll_c$. We shall first restate the definition of $\ll_c$ in a slightly different form. For the measures μ_1 and μ_2 and $\delta \in (0,1/2)$ we define functions $f_{i,\delta}$ by $f_{i,\delta}(x) = \mu_i([x-\delta,x+\delta])$. Now

(5.1) $\quad \mu_1 \ll_c \mu_2$ if and only if $f_{1,\delta}(U) \lesssim_M f_{2,\delta}(U)$ for all $\delta \in (0,1/2)$

where U is uniformly distributed on $\mathbb{C}$.

Using the assumptions in (1.6), it is straightforward to show for all $\delta \in (0,1/2)$ that $f_{1,\delta}$ is symmetric and unimodal and

$$\int_{-x}^{x} f_{1,\delta}(u)du \le \int_{-x}^{x} f_{2,\delta}(u)du \quad \text{for } 0 < x \le 1/2 .$$

For any function g on $\mathbb{C}$, let $\hat{g}$ denote its decreasing rearrangement. With m(y) equal to the Lebesgue measure of the set $\{x: g(x) > y\}$, we can define $\hat{g}(u) = \sup\{y: m(y) > u\}$ for $0 < u < 1$. Clearly

$$\int_{0}^{2x} \hat{f}_{1,\delta}(u)du = \int_{-x}^{x} f_{1,\delta}(u)du \le \int_{-x}^{x} f_{2,\delta}(u)du \le \int_{0}^{2x} \hat{f}_{2,\delta}(u)du$$

for $0 < x < 1/2$. Therefore, using a result of Hardy, Littlewood and Polya (see page 15 of Marshall and Olkin (1979)), we conclude that $f_{1,\delta}(U) \underset{M}{\le} f_{2,\delta}(U)$ as desired.

Proof of (1.7). We shall consider first the ordering $\underset{c}{\ll}$. Fix a value of b and let $X_i = D(\mu_i,b)$ for $1 \le i \le n$ and $Y = D(\mu,b)$. Then

$$D(\textstyle\sum_i p_i\mu_i, b) = \sum_i p_i X_i .$$

Let Ψ be any convex function. We know that $E\Psi(X_i) \le E\Psi(Y)$ for all i because $\mu_i \underset{c}{\ll} \mu$. Jensen's inequality now yields

$$E\Psi(\textstyle\sum_i p_i X_i) \le E\Psi(Y)$$

which completes the proof. The argument for $\underset{v}{\ll}$ is the same except that we need consider only the function $\Psi(x) = x^2$.

Proof of (1.8). The following fact is needed in the proof. Let W and Z be arbitrary bounded random variables and Ψ be any convex function having a continuous second derivative.

(5.2) $E\Psi(W) \le E\Psi((1-\beta)W + \beta Z)$ for all $\beta \in [0,1]$ if and only if $E[(Z-W)\Psi'(W)] \ge 0$.

This fact is proved by taking expectations in the inequalities

$$\Psi(W) + \beta\Psi'(W)(Z-W) \le \Psi((1-\beta)W + \beta Z) \le \Psi(W) + \beta\Psi'(W)(Z-W) + K\beta^2 .$$

The first inequality is a consequence of the convexity of Ψ. The second inequality is just the Taylor series with the remainder term replaced by an

upper bound. Let B denote the bounded set in which W and Z take values. We can choose K to be any value greater than $\frac{1}{2}\sup(\Psi''(u) : u \in B)$.

We shall prove (1.8) only for the ordering $\underset{c}{\ll}$. (The argument for $\underset{v}{\ll}$ is identical except that we consider only the function $\Psi(x) = x^2$.) Let X_i and Y be as in the proof of (1.7). We must show that $E\Psi(Y) \le E\Psi(\sum_i p_i X_i)$ for all convex functions Ψ. It suffices to verify this for smooth convex functions (having continuous second derivatives) because any convex function can be approximated arbitrarily well by a smooth one. Using (5.2) and $\mu \underset{c}{\ll} (1-\beta)\mu + \beta\mu_i$ for $0 < \beta \le 1$ we obtain $E[(X_i - Y)\Psi'(Y)] \ge 0$ for all i. Thus

$$E\left[\left((\sum_i p_i X_i) - Y\right)\Psi'(Y)\right] \ge 0$$

and applying (5.2) leads to the desired conclusion.

Proof of (2.1) is immediate from result (3.1) proved later.

Proof of (2.2). We need some preliminaries. For any vector $w = (w_1, w_2, \ldots, w_n)$ we define $F(w)$ to be the distribution on $\mathbf{R}$ which places mass $1/n$ at each point w_i so that

$$(F(w))(t) = \frac{1}{n}\sum_{i=1}^{n} I(w_i \le t).$$

It is well known (see page 17 of Marshall and Olkin (1979)) that

(5.3) $\quad v \underset{m}{\le} w$ if and only if $F(v) \underset{H}{\le} F(w)$.

Now to the proof. We shall use (2.1) and show instead that $PM(x) \underset{c}{\ll} PM(y)$ if and only if $x^{(k)} \underset{m}{\le} y^{(k)}$ for $1 \le k \le n-1$. We shall use $\overset{L}{=}$ to denote equality in distribution. It is easy to see that

(5.4) $\quad D(PM(x), k/n) \overset{L}{=} F(x^{(k)})$ for $0 \le k \le n$.

Using (5.3) and (5.4) leads to

$D(PM(x), k/n) \underset{H}{\le} D(PM(y), k/n)$ for all k if and only if $x^{(k)} \underset{m}{\le} y^{(k)}$ for all k.

It remains to show that $x^{(k)} \lesssim_m y^{(k)}$ for all k implies $D(PM(x),b) \lesssim_M D(PM(y),b)$ when $b = i/n$ for any integer i. This follows by observing that

(5.5) If $b = \lambda(i/n)+(1-\lambda)((i+1)/n)$ with $0 < \lambda < 1$ and $0 \le i \le n-1$,

then $D(PM(x),b) \overset{L}{=} \lambda F(x^{(i)})+(1-\lambda)F(x^{(i+1)})$.

Proof of (2.3). Using (1.1) we see that $PM(x) \ll PM(y)$ if and only if $ED^2(PM(x),b) \le ED^2(PM(y),b)$ for all b. From (5.4) we obtain

$$ED^2(PM(x),k/n) = \sum_{i=1}^{n}(x_i^{(k)})^2 .$$

Combining this with (5.5) leads easily to the proof.

Proof of (2.4). This result follows from (3.3) after noting that if X and Y are independent random variables on C with the same distribution μ, then the distribution of d(X,Y) is

$$\frac{n-1}{n}F((x^{(1)},x^{(2)},\ldots,x^{(n-1)})) + \frac{1}{n}F((0,1)) .$$

Proof of (3.1). We shall need some elementary facts concerning the ordering $\lesssim_M$. If the random variables X and Y satisfy EX = EY then,

(5.6) $X \lesssim_M Y$ if and only if $E(X-s)_+ \le E(Y-s)_+$ for all s,

and similarly

$X \lesssim_M Y$ if and only if $E(s-X)_+ \le E(s-Y)_+$ for all s.

See Stoyan(1983), pages 8-12. Here we have used the notation $(z)_+ = \max(z,0)$.

The result (3.1) follows immediately from the formula

(5.7) $E(D(\mu,s)-t)_+ = E(s-D(\mu^*,t))_+$.

The argument (using (5.6)) is

$\mu_1 \ll_C \mu_2$

iff $E(D(\mu_1,s)-t)_+ \le E(D(\mu_2,s)-t)_+$ for all s and t in (0,1)

iff $E(s-D(\mu_1^*,t))_+ \le E(s-D(\mu_2^*,t))_+$ for all s and t in (0,1)

iff $\mu_1^* \ll_c \mu_2^*$.

We now prove (5.7). Let $\xrightarrow{L}$ denote convergence in distribution. For distributions on $\mathbb{C}$ this has the usual meaning; $\mu_n \xrightarrow{L} \mu$ if $\mu_n(I) \to \mu(I)$ for all intervals I on $\mathbb{C}$ whose endpoints are not atoms of μ. It is straightforward to check that $\mu_n \xrightarrow{L} \mu$ implies $\mu_n^* \xrightarrow{L} \mu^*$ and $D(\mu_n,s) \xrightarrow{L} D(\mu,s)$ for all s. Thus both sides of equation (5.7) are continuous with respect to convergence in distribution. This implies that, in proving (5.7), it suffices to consider measures μ which have no atoms and have support on all of $\mathbb{C}$; any probability measure on $\mathbb{C}$ can be approximated arbitrarily well by a member of this class. In the argument which follows we restrict μ to be a member of this class.

Identify $\mathbb{C}$ with [0,1). Define F and F^* by $F(x) = \mu([0,x])$ and $F^*(x) = \mu^*([0,x])$ for all $x \in [0,1)$. The assumption on μ implies that F is invertible so that $F^{-1} = F^*$. We shall regard F and F^* as one-to-one mappings from $\mathbb{C}$ onto itself. Let V and W be independent random variables uniformly distributed on $\mathbb{C}$ and consider the random arcs [V,V+s] and [W,W+t]. Since $F^{-1} = F^*$ we have

$$(5.8) \quad P\{ [F(V),F(V+s)] \supset [W,W+t] \} = P\{ [V,V+s] \supset [F^*(W),F^*(W+t)] \} .$$

This formula is in fact true for all probability measures μ, but the argument requires more care.

Consider the left hand side of (5.8) and condition on the value of V. Clearly $P\{ [F(V),F(V+s)] \supset [W,W+t] \mid V\} = (L-t)_+$ where L is the length of the interval [F(V),F(V+s)]. Note that $L = \mu([V,V+s]) = D(\mu,s)$. This leads to

$$P\{ [F(V),F(V+s)] \supset [W,W+t] \} = E(D(\mu,s)-t)_+ .$$

If we condition on the value of W in the right hand side of (5.8) and use a similar argument we find that

$$P\{ [V,V+s] \supset [F^*(W),F^*(W+t)] \} = E(s-D(\mu^*,t))_+ .$$

This completes the proof.

Proof of (3.2). Let us reexamine the right hand side of (5.8). Define $Y = F^*(W)$. Note that Y is distributed according to μ and that $\mu([F^*(W),F^*(W+t)]) = t$. Thus, by the definition which precedes (3.2), the length of

$[F^*(W),F^*(W+t)]$ is $L(Y,\mu,t)$. This means that the right hand side of (5.8) equals $E(s-L(Y,\mu,t))_+$ so that (5.7) can be rewritten as

$$E(D(\mu,s)-t)_+ = E(s-L(Y,\mu,t))_+ \text{ where } Y \text{ is distributed according to } \mu.$$

The proof is now completed by applying (5.6) as in the proof of (3.1).

Proof of (3.3). Suppose that X and Y are independent with the same distribution μ. We first note that $Ed(X,Y)=1/2$. This is true even when μ has atoms, but in this case our convention regarding ties ($d(z,z)=0$ or 1 with equal probability) is crucial. $Ed(X,Y)=1/2$ is a consequence of $d(X,Y) \stackrel{L}{=} d(Y,X)$ and $d(X,Y)+d(Y,X)=1$ when $X \neq Y$.

Next we note that

$$ED^2(\mu,s) = 2E(s-d(X,Y))_+ \quad \text{for all } \mu. \tag{5.9}$$

We shall prove this only for μ which have no atoms and have support on all of $\mathbb{C}$. (The proof for general μ can then be obtained by a limiting argument.) Go back to (5.8) and replace t by a random variable U which is uniformly distributed on [0,1) and independent of V and W. Now W and W+U are independent and uniformly distributed on $\mathbb{C}$ so that $F^*(W)$ and $F^*(W+U)$ are independent with the distribution μ. The length of $[F^*(W),F^*(W+U)]$ is thus $d(X,Y)$ with $X = F^*(W)$ and $Y = F^*(W+U)$. Repeating the argument in the proof of (5.7) now yields $E(D(\mu,s)-U)_+ = E(s-d(X,Y))_+$. Because

$$\int_0^1 (x-u)_+ du = x^2/2 \quad \text{for } 0 \le x \le 1$$

we have $E(D(\mu,s)-U)_+ = (ED^2(\mu,s))/2$ and the proof of (5.9) is complete.

Using (5.9) we see that $ED^2(\mu_1,s) \le ED^2(\mu_2,s)$ for all s if and only if $E(s-d(X_1,Y_1))_+ \le E(s-d(X_2,Y_2))_+$ for all s. Now (1.1) and (5.6) lead us immediately to (3.3).

Proof of (4.1). First note that $E\mu_1(\mathcal{A}) = E\mu_2(\mathcal{A})$. This follows from Fubini's theorem which gives

$$E\mu(\mathcal{A}) = \int_{\mathbb{C}} P(x \in \mathcal{A}) d\mu(x)$$

and the fact that the distribution of the random set $\mathcal{A}$ is rotationally invariant. From the equality of means we see immediately that $\mathrm{Var}(\mu_1(\mathcal{A})) \leq \mathrm{Var}(\mu_2(\mathcal{A}))$ if and only if $E(1-\mu_1(\mathcal{A}))^2 \leq E(1-\mu_2(\mathcal{A}))^2$.

Using $1-\mu(\mathcal{A}) = \mu(\mathcal{A}^c)$ (with c denoting complement) and Fubini's theorem we obtain

$$E(1-\mu(\mathcal{A}))^2 = \int_C \int_C P(x \in \mathcal{A}^c, y \in \mathcal{A}^c) d\mu(x) d\mu(y) . \tag{5.10}$$

Define $\mathbb{B}_i = [U_i, U_i+b_i)$ so that $\mathcal{A}$ is the union of $\mathbb{B}_1, \mathbb{B}_2, \ldots, \mathbb{B}_n$. The intervals $\mathbb{B}_i$ are independent so that

$$P(x \in \mathcal{A}^c, y \in \mathcal{A}^c) = \prod_{i=1}^{n} P(x \in \mathbb{B}_i^c, y \in \mathbb{B}_i^c)$$

where it is easily seen that

$$P(x \in \mathbb{B}_i^c, y \in \mathbb{B}_i^c) = (d(x,y)-b_i)_+ + (1-d(x,y)-b_i)_+ .$$

Thus (5.10) can be rewritten as $E(1-\mu(\mathcal{A}))^2 = E\Psi(d(X,Y))$ where X and Y are independent with distribution μ and

$$\Psi(z) = \prod_{i=1}^{n} [(z-b_i)_+ + (1-z-b_i)_+] .$$

It is straightforward to show that Ψ is convex so that (4.1) now follows from result (3.3).

Proof of (4.2). To simplify the presentation we carry out the proof only for measures μ_3 which are discrete with finitely many atoms. Suppose μ_3 has atoms of mass p_i at the points x_i for $1 \leq i \leq n$. Then

$$\mu_1(\mathcal{A}) = \sum_i p_i \mu_2(\mathcal{A} - x_i)$$

so that

$$\Psi(\mu_1(\mathcal{A})) \leq \sum_i p_i \Psi(\mu_2(\mathcal{A} - x_i))$$

holds for any convex function Ψ. Now taking expectations and using the rotational invariance of the distribution of $\mathcal{A}$ leads to $E\Psi(\mu_1(\mathcal{A})) \leq E\Psi(\mu_2(\mathcal{A}))$ as desired.

Proof of (4.3). We shall prove this only for μ_1, μ_2 which have no atoms and have support on all of $\mathbb{C}$. Let $U_1, U_2, \ldots, U_n$ be i.i.d. uniform on $\mathbb{C}$. Suppose $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$ are the points U_i arranged in counterclockwise order beginning with $U_1 = U_{(1)}$. Identify $\mathbb{C}$ with $[0,1)$ and define F_i by $F_i(x) = \mu_i([0,x))$ for $i = 1, 2$. Now define the dual meaures μ_1^* and μ_2^* on $\mathbb{C}$ by $\mu_i^*([0,x)) = F_i^{-1}(x)$. The points $F_i^{-1}(U_1), F_i^{-1}(U_2), \ldots, F_i^{-1}(U_n)$ are i.i.d. from the distribution μ_i so that we have

$$\sum_{i=1}^{k} S_i \overset{L}{=} \mu_1^*([U_{(1)}, U_{(k+1)})) \text{ and } \sum_{i=1}^{k} T_i \overset{L}{=} \mu_2^*([U_{(1)}, U_{(k+1)})).$$

Let $\mathcal{B} = [U_{(1)}, U_{(k+1)})$. Conditional on its length, $\mathcal{B}$ is just a uniformly placed random arc on $\mathbb{C}$ so that $\mu_1^*(\mathcal{B}) \overset{st}{\leq} \mu_2^*(\mathcal{B})$ follows immediately from (3.1). This completes the proof.

Statistics for testing uniformity. We now show that the scan statistic and Kuiper's statistic are c-increasing. Let $s = (s_1, s_2, \ldots, s_n)$ be the vector of consecutive spacings between n points on $\mathbb{C}$ and let $N(x,b)$ denote the number of points on $\mathbb{C}$ contained in the interval $(x, x+b)$. The scan statistic $N(b)$ having width parameter b is defined as $N(b) = \sup_x N(x,b)$ with x ranging over $\mathbb{C}$. The scan statistic has an obvious relationship with the higher order spacings;

$$N(b) \geq k+1 \text{ if and only if } \min_i s_i^{(k)} < b$$
$$\text{if and only if } \max_i s_i^{(n-k)} > 1-b .$$

The function $\Phi(x_1, x_2, \ldots, x_n) = \max_i x_i$ is Schur-convex so that (4.4) implies the function $\phi(s) = \max_i s_i^{(n-k)}$ is c-increasing which in turn implies that the scan statistic is c-increasing.

Now we consider Kuiper's test. Choose an arbitrary zero point on C and identify C with the interval [0,1). Let $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$ be the order statistics of the n points in our sample. Kuiper's test statistic V_n may be computed using the formula

$$V_n = \frac{1}{n} + \max_i \left(y_{(i)} - \frac{i}{n}\right) - \min_i \left(y_{(i)} - \frac{i}{n}\right).$$

See page 174 of Mardia(1972). By arguing as in Section 4 of Cressie (1977), it is straightforward to reexpress Kuiper's test in terms of the spacings as

$$V_n = \frac{1}{n} + \max_k \left\{ \max_i \left(s - \frac{1}{n}\right)_i^{(k)} \right\} \tag{5.11}$$

where $s - \frac{1}{n}$ denotes the vector $(s_1 - \frac{1}{n}, s_2 - \frac{1}{n}, \ldots, s_n - \frac{1}{n})$. The function inside braces in (5.11) is c-increasing by (4.4). Now, since the maximum of c-increasing functions is again c-increasing, we see that V_n is c-increasing as desired.

References

Beran, R.J. (1968). Testing for uniformity on a compact homogeneous space. *J. Appl. Prob.* **5**, 177-95.

Beran, R.J. (1969). Asymptotic theory of a class of tests for uniformity of a circular distribution. *Ann. Math. Statist.* **40**, 1196-206.

Birnbaum, Z.W. (1948). On random variables with comparable peakedness. *Ann. Math. Statist.* **19**, 76-81.

Cressie, N. (1976). On the logarithms of high-order spacings. *Biometrika* **63**, 343-55.

Cressie, N. (1977). On some properties of the scan statistic on the circle and the line. *J. Appl. Prob.* **14**, 272-83.

Cressie, N. (1979). An optimal statistic based on higher order gaps. *Biometrika* **66**, 619-27.

Guttorp, P. and R. A. Lockhart (1988). On the asymptotic distribution of quadratic forms in uniform order statistics. *Ann. Statist.* **16**, 433-49.

Hardy, G.H., J.E. Littlewood and G. Polya (1952). *Inequalities*, 2nd ed. Cambridge Univ. Press, London and New York.

Holst, L. (1979). Asymptotic normality of sum-functions of spacings. *Ann. Probab.* **7**, 1066-72.

Mardia, K.V. (1972). *Statistics of Directional Data*. Academic Press, New York.

Marshall, A.W. and I. Olkin (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

Naus, J. I. (1966). A power comparison of two tests of non-random clustering. *Technometrics* **8**, 493-517.

Pyke, R. (1965). Spacings (with discussion). *J. Roy. Statist. Soc. Ser. B* **7**, 395-449.

Stoyan, D. and (edited with revisions by) D.J. Daley (1983). *Comparison Methods for Queues and Other Stochastic Models.* Wiley, New York.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 406	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Ordering Distributions On The Circle With Respect To Uniformity		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Fred W. Huffer		8. CONTRACT OR GRANT NUMBER(s) N00014-86-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE June 29, 1988 13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Also issued as AFOSR Technical Report No. 88-228, Florida State University.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

spacings, random arcs, tests of unformity, directional data, majorization, convex ordering.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE FOLLOWING PAGE.

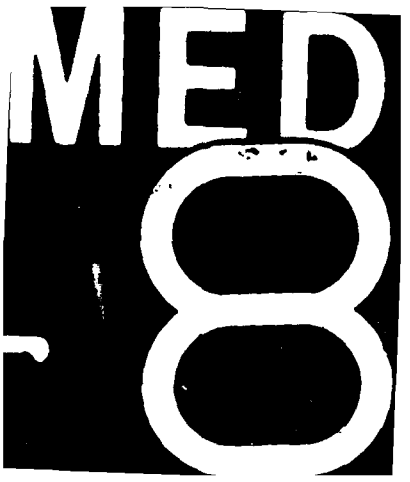
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TECHNICAL REPORT NO. 406

20. ABSTRACT

Orderings (denoted <<) for probability distributions on the circle are introduced for which $\mu_1 << \mu_2$ means roughly that μ_1 is "more uniform" than μ_2, or that μ_2 is "more clumped" than μ_1. Somewhat more precisely, $\mu_1 << \mu_2$ if the random variable $\mu_1(A_s)$ is "less variable" than the random variable $\mu_2(A_s)$ for all s where A_s denotes a random arc of length s distributed uniformly on the circle. Some properties of the orderings are explored and applications are presented to random coverage problems on the circle, the distribution of spacings and the analysis of directional data.


MED
8